Impact of the timing of a SAR image acquisition on the calibration of a flood inundation model

Sacha Gobeyn[a], Alexandra Van Wesemael[a,*], Jeffrey Neal[b], Hans Lievens[a], Katrien Van Eerdenbrugh[a], Niels De Vleeschouwer[a], Hilde Vernieuwe[c], Guy J.-P. Schumann[b,d], Giuliano Di Baldassarre[e], Bernard De Baets[c], Paul D. Bates[b], Niko E. C. Verhoest[a]

[a]*Laboratory of Hydrology and Water Management, Ghent University, Coupure links 653, 9000 Ghent, Belgium*
[b]*School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1SS, UK*
[c]*KERMIT, Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Coupure links 653, 9000 Ghent, Belgium*
[d]*Remote Sensing Solutions, Inc., 248 E Foothill Blvd, Suite 200, Monrovia CA 91016, USA*
[e]*Department of Earth Sciences, Uppsala University, Villav. 16, 75236 Uppsala, Sweden*

Abstract

Synthetic Aperture Radar (SAR) data have proven to be a very useful source of information for the calibration of flood inundation models. Previous studies have focused on assigning uncertainties to SAR images in order to improve flood forecast systems (e.g. Giustarini et al. (2015) and Stephens et al. (2012)). This paper investigates whether the timing of a SAR acquisition of a flood has an important impact on the calibration of a flood inundation model. As no suitable time series of SAR data exists, we generate a sequence of consistent SAR images through the use of a synthetic framework. This framework uses two available ERS-2 SAR images of the study area, one taken during the flood event of interest, the second taken during a dry reference period. The obtained synthetic observations at different points in time during the flood event are used to calibrate the flood inundation model. The results of this study indicate that the uncertainty of the roughness parameters is lower when the model is calibrated with an image taken before rather than during or after the flood

*Corresponding author
Email address: Alexandra.VanWesemael@UGent.be
URL: http://www.ugent.be/bw/dfwm/en/research/LHWM/

peak. The results also show that the error on the modelled extent is much lower when the model is calibrated with a pre-flood peak image than when calibrated with a near-flood peak or a post-flood peak image. It is concluded that the timing of the SAR image acquisition of the flood has a clear impact on the model calibration and consequently on the precision of the predicted flood extent.

Keywords:
flood inundation, hydraulic modelling, remote sensing, calibration

---

1. Introduction

Hydrodynamic models have become an essential tool for the prediction of inundation events. These mathematical models, hereafter referred to as flood inundation models, traditionally require the channel and floodplain Manning coefficients to be calibrated with *in situ* measurements (e.g. water height, flow) from gauging stations at the downstream boundary of the model (Bates et al., 1992). For example, assuming a constant Manning coefficient over the channel and floodplain, one can calibrate both parameters by minimizing e.g. the root mean square error (RMSE) between the time series of the observed and simulated water height (or flow). This approach only evaluates the simulation of bulk flow routing to that specific point, hence the model dynamics in the floodplain is not evaluated (Bates et al., 2004). However, the variable of interest for most end-users is the water level at different places in the floodplain, or a product derived from this, e.g. the flood extent (Hunter et al., 2007). In order to evaluate the inundation predicted by a model, the simulations should be validated with spatially-distributed water level data or flood extent data. Among alternatives to evaluate the model's performance distributedly, are the use of post-flood field survey data (Aronica et al., 1998; Hunter et al., 2005; Neal et al., 2009), aerial photos or airborne and spaceborne Synthetic Aperture Radar (SAR) data (Bates and De Roo, 2000; Aronica et al., 2002; Horritt et al., 2007; Di Baldassarre et al., 2009a). Especially, radar, with its day, night and

cloud penetrating capacity, is a promising technology for supporting flood inundation modelling (Bates, 2004; Montanari et al., 2009; Schumann et al., 2009b; Tarpanelli et al., 2012).

Many aspects of calibrating flood inundation models with SAR are described in the scientific literature. One aspect concerns the fact that the parameter space is often restricted to the Manning coefficients as these parameters are considered to be the most important parameters controlling the flow characteristics (Romanowicz et al., 1996), even though it is recognized that several parameters (related to roughness, channel cross section dimensions, initial conditions) influence the model results. Another aspect involves the use of different performance measures in the calibration process. The performance measures presented by Aronica et al. (2002) and Hunter et al. (2005) are the measures most commonly used to evaluate flood inundation models with SAR-extracted binary flood maps. Essentially, these measures are based on the confusion matrix. As a last aspect, it is recognized that the performance of the calibration depends on the accuracy of the extracted flood map. The extraction of a flood from a SAR image typically results from a delineation algorithm, which defines pixels as being flooded based on their backscatter value. Yet, such exercise is prone to uncertainty (Matgen et al., 2011; Schumann et al., 2014; Giustarini et al., 2015). Some good examples of delineation algorithms can be found in the studies of Horritt (1999), Mason et al. (2007), Schumann et al. (2009a) and Giustarini et al. (2013). Schumann et al. (2014) employed a slightly different approach by avoiding the need for an *a priori* classification of the SAR image by calibrating the roughness parameters on an ensemble of SAR-extracted binary maps.

In many studies, models were calibrated using spaceborne SAR imagery (e.g. Tarpanelli et al. (2012)), however, the link between the calibration results and the effect of the timing of the SAR image acquisition with respect to the flood peak has not been explicitly addressed. For instance, García-Pintado et al. (2013) found that simulated ensembles of water height showed the largest deviation during the rising limb

of the storm event analysed, leading to an increased performance of the assimilation of SAR-derived water level observations into hydrodynamic models. A second study, presented by Horritt et al. (2007), indicated that there is a possible effect of the timing of the image acquisition on the calibration. Horritt et al. (2007) were unable to properly constrain the optimal values for the LISFLOOD-FP model (Bates and De Roo, 2000) when using SAR images taken near peak flow as calibration data. This problem of being unable to properly constrain the parameter space is closely linked to the equifinality problem, *i.e.* different parameter sets yield similar model results with respect to the objective function used in the calibration (Beven and Binley, 1992; Aronica et al., 1998, 2002; Horritt and Bates, 2002; Pappenberger et al., 2005; Werner et al., 2005). The parameter sets that lead to the same optimal solution for the calibration problem are often defined as 'behavioural' parameter sets (Beven, 1996). Very often, a large number of behavioural parameter sets are identified when the floodplain Manning coefficient is included in the calibration process. Aronica et al. (2002) found that the 1D/2D-hybrid model LISFLOOD-FP (Bates and De Roo, 2000) showed a lower sensitivity to the floodplain Manning coefficient than towards the channel Manning coefficient while Horritt and Bates (2002) came to a similar conclusion for the LISFLOOD-FP model, they found a higher sensitivity to floodplain roughness values for the finite element model TELEMAC-2D and the 1D model HEC-RAS.

In this paper, we assess the impact of the timing of a flood observation with respect to the peak flow on the calibration of a flood inundation model. We expect that roughness parameters will mainly determine the rate of flooding and receding, and thus having observations before or after the flood peak should allow one to discriminate better between roughness parameters than when an observation is taken near the maximum flood extent. Therefore, we restrict the calibration to only the roughness parameters (e.g. as in Romanowicz et al. (1996); Aronica et al. (1998); Horritt and Bates (2001); Hostache et al. (2009); Di Baldassarre et al. (2009a); Schumann et al. (2014)). We focus on the River Dee, which has been used several times to test different hypotheses and techniques

related to flood inundation modelling (Di Baldassarre et al., 2009a; Stephens et al., 2012; Schumann et al., 2014). ~~One SAR image of the inundation event is available. To assess the impact of the timing of the image acquisition on the calibration, we need multiple and consistent images for experimental control.~~

Section 2 describes the study site, the model, the calibration method and the synthetic framework. Section 3.1 presents the calibration with the available SAR data. Section 3.2 discusses the results of the synthetic experiment in which we test whether the timing of the SAR image acquisition has an influence on the calibration. Section 3.4 elaborates on the uncertainty of the simulated flood extent based on the calibration results. Finally, Section 4 presents the conclusions for this work.

2. Methodology

2.1. Study site and test data

The River Dee originates in the Snowdonia region in North Wales and flows approximately 110 km to its mouth north of the city of Chester where it flows into Liverpool Bay. The catchment area covers 1816.8 km^2 and the mean observed flow at the downstream end of the study site, is about 30 m^3/s. The selected area for this case study is located at the confluence of the River Dee and the River Alyn (Figure 1), an area with mainly agricultural activity which has shown to be prone to severe floods. The total modelled area is about 40 km^2 and covers an 11-km section of the River Dee, between Farndon and Eaton Hall, and 9 km of the River Alyn. The River Alyn is on average 12 m wide, while the River Dee has a mean width of 30 m. The River Dee flows in a northerly direction, with the floodplain only about 150 m in width for the first few kilometers of the study area. After this, the floodplain reaches a width of 2 km, narrowing to 0.5 km for the final 3 km of the study domain. The lower part of the river is subject to tidal effects, which can affect the river levels at Ironbridge, 15 km upstream of Chester Weir. The combination of these tidal effects and high magnitude precipitation events has caused multiple floods over the considered

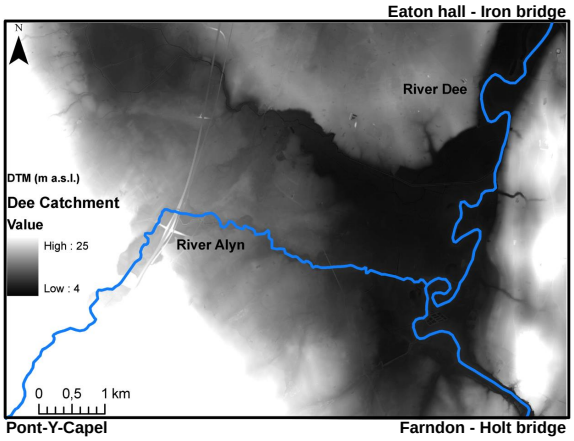


Figure 1: Illustration of the height profile of the study area. Pont-y-Capel and Holt bridge are the upper boundaries of the study area, while Ironbridge is the downstream boundary.

reach (Di Baldassarre et al., 2009a; Stephens et al., 2012). The study site was chosen for its availability of a satellite image of a~~n extensive~~ 2-year flood event that occurred in December 2006. Information about the height profile of the floodplain is available through LiDAR data (2 m resolution, vertical accuracy of 0.15 m) and Interferometric Synthetic Aperture Radar (InSAR) data (5 m resolution, vertical accuracy of 1 m). Gauged stage **and/or flow data are** provided upstream for the River Alyn at Pont-y-Capel, upstream for the River Dee at Holt bridge at Farndon and downstream for the River Dee at Ironbridge at Eaton Hall **at a time step of 15 minutes**. ~~Furthermore, gauged flow data are provided upstream at Pont-y-Capel and downstream at Ironbridge, while the stage data at Holt bridge are converted to flows using a flow rating curve assuming mass conservation at the confluence (Stephens et al.,2012). All data are available at a time step of 15 minutes~~

During the flood on December 12, 2006, the satellite sensor ERS-2 (VV-polarisation in a descending track) observed the inundation, with a pixel spacing of 12.5 m, when the flood was receding, at 11:07 am (Schumann et al., 2009b). **For further analysis, pixel spacing was changed to 20 m using a linear**

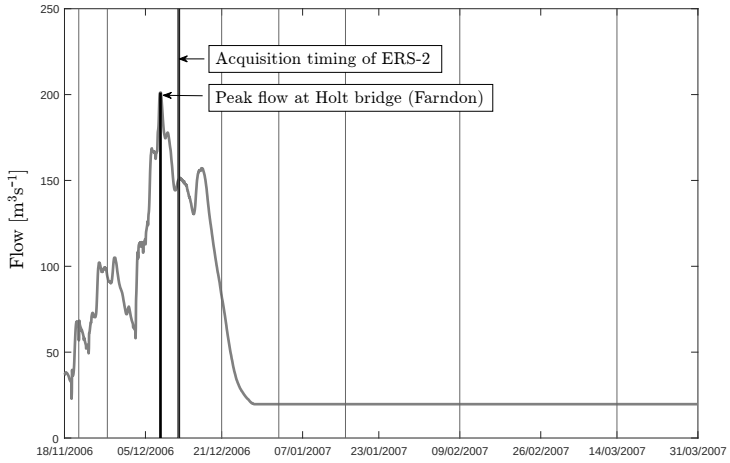


Figure 2: Illustration of the inflow hydrograph at Holt bridge (Farndon). The satellite observed the flood on December 12, 2006, 11:00 am. Peak flow is observed on **December 8 at 2:15 pm. Points in time of the synthetic observations are also indicated with gray lines**. After the storm has receded, the hydrograph is kept at a stable level in order to analyse an isolated storm event.

interpolation method. The inflow hydrograph for the River Dee (at Holt bridge, Farndon) and the acquisition time of the ERS-2 image are illustrated in Figure 2. As only one storm event is analysed in this study, the hydrograph is kept at a stable level after recession of the river flow in order to exclude any influences from new upcoming storm events.

2.2. Flood inundation modelling

In this study, the LISFLOOD-FP model is used to simulate the flood dynamics and extent of the December 2006 event. This 1D/2D hybrid model allows for the simulation of a flood event by coupling channel dynamics, defined as a one-dimensional vector, to floodplain dynamics, defined over a raster grid (Bates and De Roo, 2000). In this paper, we use a specific formulation of the LISFLOOD-FP model developed by Neal et al. (2012). This two-dimensional floodplain model uses an explicit forward difference scheme to simulate a shallow water wave (without convective acceleration) over the floodplain **(Bates et al.,**

2010). The channelized flows are simulated by a subgrid scale representation, defined by projecting the cross section properties of the river onto the 1D river network layer. In this study, the cross sections are defined as rectangular sections with the channel bed elevation and width parameterized based on ground surveyed river cross sections. This simplification of the cross sections may be justified for channels that are wide and shallow with cohesive banks. Furthermore, it makes the model simple while the shape of the channel cross section is less important for out of bank flows. This rectangular cross section assumption therefore can be considered as a reasonable approximation.

The Manning (roughness) coefficient is, in principle, a spatio-temporally-distributed parameter as it changes over space (e.g. bed channel versus vegetated floodplains) and time (e.g. seasonal changes of vegetation). It theoretically depends on the water level, *i.e.* friction tends to decrease for higher water levels. However, such a spatio-temporal variability of the Manning coefficient cannot be measured in real world applications (Di Baldassarre, 2012). Moreover, Manning coefficients are generally determined under conditions of uniform flow, which are rarely observed in nature. Thus, usually, in flood inundation modelling, the Manning coefficients are regarded as free parameters to be estimated via model calibration (Anderson et al., 1996). However, the use of numerous Manning coefficients within the model can lead to over-fitting, *i.e.* the model performs well in calibration as more parameters improve the fit, but miserably fail in predicting other events than the calibration event. Thus, state-of-the-art methods for flood inundation modelling (Horritt and Bates, 2001; Pappenberger et al., 2006; Horritt et al., 2007; Di Baldassarre et al., 2009a; Neal et al., 2013) only distinguish between Manning coefficients in the main channel and floodplain and assume they are temporally and spatially constant. In this study, we follow the same assumption and thus only discern between a Manning coefficient for the floodplain (n_{fp}) and for the channel (n_{ch}).

The advantage of the model is that it is able to operate with low- and high-resolution digital elevation models, and, furthermore, it is straightforward to be validated with spatially-distributed raster data (e.g. SAR data). For the

specific mathematical description of the model we refer to Bates et al. (2010) and Neal et al. (2012).

The area of interest contains three boundary conditions. The downstream boundary is defined at Ironbridge and stage data were implemented as such. The river Alyn, upstream of its confluence with the river Dee, seemed to have little influence on the flooding. Since also flow variation here was minimal, a stable flow level was used. For the second upstream boundary at Holt bridge, only water level data were available. In order to generate input flow data, mass conservation at the confluence of the river Dee and the river Alyn was performed, using flow measurements from both other gauging stations (Stephens et al., 2012). So lastly, for the Holt bridge boundary, an inflow hydrograph was generated. The grid resolution of the model is 20 m, in order to have an acceptable run-time for the simulation (90 minutes on an Intel® Core™2 Quad Processor, 2.66 GHz), while not losing too much detail for representing the inundation dynamics (Horritt and Bates, 2001). A LiDAR-derived Digital Elevation Model (DEM) (Stephens et al., 2012) was used as a basis for the 20-m grid. Some small changes to the DEM have been made in order to account for the following concern. In the LISFLOOD-FP model, floodplain flow is a function of the free surface slope. As micro-scale slopes are not always captured in a 20 m DEM resolution causing that the natural routing network **could unintentionally be ignored**. This causes flood water to remain in structural depressions for a prolonged time within the study area. Therefore, a small drainage network has been **burned onto** the DEM. It is important to note that results, especially concerning the recession of the flood, can be affected to some extent, although it is expected that main trends should remain the same. Furthermore, the sub-grid channel elements were determined through interpolating and projecting measured channel geometries (provided by the Environment Agency of England and Wales) onto the river network layer. ~~Additionally, the acquired channel geometries were tested and validated using OS Digimap data *(Crown Copyright/database right 2014. An Ordnance Survey/EDINA supplied service)*~~

2.3. ROC-based calibration

The receiver operating characteristics (ROC) method (Schumann et al., 2014) is used to calibrate the channel and floodplain Manning coefficients. The ROC graph is a method to visualize, organize and select classifiers based on their performance (Fawcett, 2006). The adopted performance index, the area under the ROC curve (AUC), is often employed to evaluate the performance of different classifiers (Bradley, 1997). The proof-of-concept of this alternative approach to calibrate flood models with spatially-distributed data was presented by Schumann et al. (2014), who employed the idea that a ROC curve reflects the ability of a classification method to distinguish between two classes (in this case flooded and non-flooded areas). In the calibration of a flood model this is performed as follows. For each model parameter set that is tested, the corresponding binary flood map is established and compared with the original SAR image in terms of the AUC. By doing so, all information present in the SAR image is employed in order to select the model parameter set that yields the best representation of this information in a modelled binary flood map, *i.e.* the parameter set that results in the highest AUC. Schumann et al. (2014) advocate the use of this method over the normally-used methods in conjunction with binary flood maps as this method does not rely on an *a priori* classification of the images that is prone to uncertainty (Matgen et al., 2011).

The approach of identifying the ROC and its AUC is briefly explained as follows: for one threshold of the SAR backscatter value (expressed in decibel, dB), the delineated observed flood map, *i.e.* flooded pixels are identified when their backscatter values are lower than the chosen threshold, is used together with the modelled binary flood map to determine the confusion matrix. Based on this confusion matrix, the true positive rate (T_p), *i.e.* the proportion of positive (flooded) instances correctly classified as positive (flooded), is plotted versus the false positive rate (F_p), *i.e.* the proportion of the negative instances (non-flooded) erroneously classified as positive (flooded). By moving the threshold of the SAR backscatter value from the minimum to the maximum backscatter value within the SAR image, different confusion matrices and thus

different points (T_p,F_p) in the ROC space can be determined that hence constitute the ROC curve. For different parameters of the flood inundation model, different binary flood maps are obtained for which different ROC curves can be determined. **This method of thresholding assumes that water surfaces always generate low backscatter values, inevitably introducing errors in the pattern of the flood. Those errors can be due to shadow effects on the SAR image being classified as flooded area, while vegetation on the other hand can hamper the detection of flooded pixels, especially near the boundaries of the flood. Many other delineation algorithms exist, however, thresholding is above all computationally efficient.**

Based on a wide range of possible Manning coefficient values reported by Chow et al. (1988), the intervals of 0.02-0.10 and 0.04-0.16 are chosen respectively for the channel and floodplain Manning coefficient, respectively. In this exercise, the parameter space is explored over a fixed grid in steps of 0.005 for the channel and 0.02 for the floodplain Manning coefficient. **Step size for both Manning coefficients is not equal, and based on preliminary research results and expert knowledge.** The ROC curve and the AUC, ranging from 0 to 1, is determined for 119 grid points (*i.e.* 17 for the channel Manning coefficient, 7 for the floodplain Manning coefficient). The highest AUC is compared with the AUCs of the other 118 points with the pairwise comparison test of DeLong et al. (1988) (with Bonferoni-Holm correction on the p-values). The parameter set with the highest AUC is considered as a behavioural set, as well as all other sets that do not have an AUC value that is significantly different from this highest AUC value. The number of behavioural sets N is used as a measure for the ability of the data to constrain the parameter space of the model.

2.4. Synthetic framework

To assess the impact of the timing of a SAR image acquisition on the calibration of the model, a time series of SAR images during the flood event is needed. More in particular, for experimental control, we need multiple and consistent

SAR images representing a single flood event. **Currently, the Sentinel constellation (Sentinel-1A and B) together with the COSMO-SkyMed mission should be able to constitute a multi-temporal series of SAR images capturing a flood event. However, as this was not yet the case at the time of our case study, a framework** is set up that allows for the generation of synthetic SAR images at different instants during the flood event using the backscatter information of an existing SAR image. For this case study, one ERS-2 radar image of the inundation event is available that was acquired on December 12, 2006, 3 days after the peak flow registered at Holt bridge. The framework used to perform the calibration with synthetic observations is illustrated in Figure 3 and is based on an *identical twin experiment*. In such experiment, for each chosen point in time, the 'true state' of a system is generated by a model (*i.e.* LISFLOOD-FP), while the simulation of this true state is generated by the same model. In this study, the true state is a binary representation of the flood, which does not represent the real system, but only a hypothetical state in which the system could be (Matgen et al., 2010). This true state is then converted into a synthetic ~~SAR~~ observation by **a method of choice** ~~applying a bootstrap method (Efron, 1979) using the ERS-2 SAR image and a non-flooded reference SAR image that was taken over the same site on August 29, 2006~~. The generation of the true state is explained in Section 2.4.1 and used in Section 2.4.2 where the signal characteristics of both the flooded SAR image and the dry reference image are combined to generate the 'synthetic scene'. This processed image serves as a synthetic observation to perform the ROC-based calibration with, for that considered point in time. The objective is to find N behavioural sets $\theta_{opt,i}$ ($i \in \{1, .., N\}$) that can be compared with the assumed true parameter set $\theta_{opt,true}$, which is used to generate the synthetic true state. This process can be repeated for several points in time. For this case study, the results are compared for **synthetic** images on 8 points in time between November 21 and March 14, 2006, **as indicated in figure 2**. In this way, more data are obtained than **what were actually available** ~~could possibly be obtained from any current SAR sensor~~, which justifies the use of

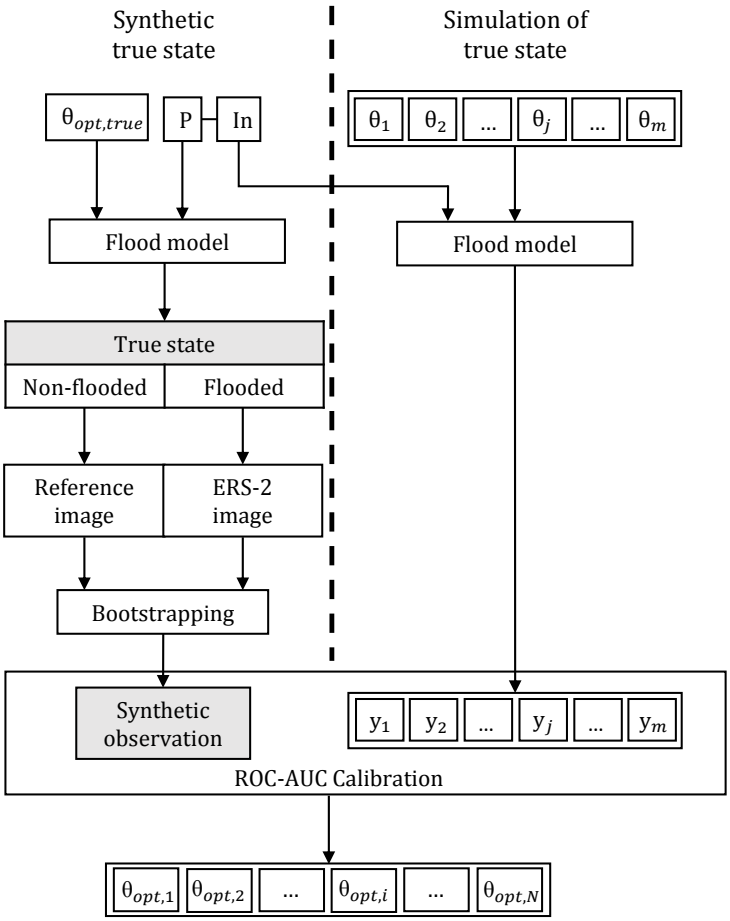


Figure 3: Synthetic framework. $\theta_{opt,true}$ is the true parameter set, In represents the input, P refers to the perturbation function defined in Eq.(1), y_j is the j^{th} model output corresponding to parameter set θ_j. $\theta_{opt,i}$ is the i^{th} behavioural parameter set found through calibration. The behavioural sets found, are compared with the true set.

our synthetic framework.

2.4.1. Generation of the true state

The goal of this section is to generate the true state for a certain point in time, which is a binary map of the flood, by using a model simulation. In order to take into account errors and uncertainties, we chose to perturb the input of the flood inundation model. The errors in the model structure are not accounted for, because the assumption is made that errors in model state and output are mainly caused by errors in input (Pappenberger et al., 2006; Matgen et al.,

2010; García-Pintado et al., 2013). The true state of the system is generated by propagating a perturbation of the input flow through the flood inundation model, using an autoregressive model:

$$q_j = \alpha q_{j-1} + \sqrt{1-\alpha^2} w_j \tag{1}$$

where q_j (m^3/s) is the input flow error at point in time j, α (-) the temporal autocorrelation coefficient (Evensen, 2003) and w_j (m^3/s) a white noise term obtained by drawing random samples from a normal distribution $\mathcal{N}(0, \sigma^2)$. The flow error is assumed to be a multiplicative error, *i.e.* larger flows have a higher error rate (Di Baldassarre and Montanari, 2009). A value of 15% is assumed for σ as an overall error affecting the river flow observations, based on the errors on flow measurements reported by Di Baldassarre and Montanari (2009); Pappenberger et al. (2006) and Lang et al. (2010). The temporal autocorrelation coefficient in Eq. (1) is represented by an exponential decay (García-Pintado et al., 2013):

$$\alpha = e^{-\frac{\Delta t_j}{\tau}} \tag{2}$$

where Δt_j (s) is the length of the time interval between points in time $j-1$ and j and τ (s) the decorrelation length. The specific point in time j is known, however, the decorrelation length should be determined or estimated. For the choice of the decorrelation length, it is assumed that the temporal autocorrelation of the flow error is equal to that of the flow. A decorrelation length of 4 days was chosen by analysing the stability of the perturbed flow time series, since large fluctuations in the input flow time series may introduce calculation instabilities in the model. In addition, other values for the decorrelation length were analysed, however, lower values resulted in unstable inflow curves and thus a very steep descending autocorrelation function of the perturbed inflow time series. On the other hand, higher values for the decorrelation length resulted in simply adding a bias to the inflow time series.

2.4.2. Generation of synthetic observations

The main goal of the synthetic framework is to produce **a time series of** ERS-2 SAR-like images for each chosen point in time that will serve as observational data for the calibration method. ~~As stated before, with current techniques, a high-frequency time series of consistent SAR images are not available, requiring that we produce synthetic SAR images if we wish to assess the impact of timing of a SAR image acquisition on the calibration of the flood inundation model.~~ To do so, the LISFLOOD-FP-generated true state binary maps are used as the base layer to create scene images. For this purpose, a bootstrap method (Efron, 1979) is performed using the flooded ERS-2 SAR image (December 2006) and **a** dry reference ERS-2 SAR image **that was taken over the same site on August 29, 2006.** The latter, however, first requires some image pre-processing to be consistent with the flooded image. First, a Sigma Lee filter with a 9×9 window (similar to Schumann et al. (2009a)) is applied for speckle filtering. Then, to eliminate seasonal differences and to take into account different wetness conditions between both figures, the non-flooded pixels of the reference image are rescaled by quantile mapping (or cdf-matching) to the non-flooded pixels in the ERS-2 SAR image. To distinguish flooded from non-flooded pixels in the reference image, the delineation provided in Schumann et al. (2009b) is used. **This delineation is based on mainly visual processing and is further supported by expert knowledge. The** pre-processing steps **mentioned**, ensure that the rescaled non-flooded image has similar characteristics as of an image taken during the flooding period. For the true state binary maps, the pixels that are being considered as flooded, are assigned backscatter values by bootstrapping flooded pixels in the ERS-2 SAR image. Pixels that are being considered as non-flooded on the other hand, are bootstrapped from the rescaled reference image. However, to ensure that the study area preserves the spatial pattern in the SAR image, the bootstrapped values are assigned to flooded (non-flooded) pixels in such a way that all flooded (non-flooded) pixels keep the same ranking order as those in the reference image. This approach is

similar to Giustarini et al. (2015) and ensures that the resulting synthetic SAR image is similar to the original speckle filtered ERS-2 image.

2.5. Assessment of the synthetic observations

The histogram of the backscatter values of a flood SAR image typically has a bimodal shape, with the first peak corresponding to the open water pixels and the second peak corresponding to the non-flooded pixels (Matgen et al., 2011). The same should thus be encountered for the synthetic images created in Section 2.4. The histogram of these backscatter values should also follow this typical bimodal distribution. In order to validate the approach used in Section 2.4 (see Figure 4), the synthetic image of December 12, 2006, 11:07 am is compared with the actual ERS-2 SAR image, acquired at the same point in time. As can be seen in the bottom panel of Figure 4, the empirical probability distribution **function** (pdf) of the synthetic observation is clearly bimodal. Generally, the synthetic image follows the same type of distribution as the original image, **although the valley between both peaks of the latter is less profound** ~~pdf of the original image is less pronounced~~. Two peaks can clearly be distinguished (around -8 and -10 dB) with a minimum in between at -9 dB. Mainly the shape of the pdf in the valley between both peaks should be reproduced well, as this will determine whether both images may discern similarly between flooded and non-flooded pixels. The non-parametric Kolmogorov-Smirnov test states that at the 5% significance level the null hypothesis that both histograms of backscatter values originate from the same distribution, is accepted. Visual inspection of the top panel in Figure 4 shows that the edges between flooded and non-flooded areas are less fuzzy compared to the edges in the original ERS-2 image. This deficiency is probably due to the fact that the bootstrap sample of the flooded pixels, which was digitized from the original SAR image, does not contain sufficient pixels within the transition zone between the flooded and non-flooded area. **Although this is the area where uncertainty mainly appears due to low vegetation hampering the clear detection of flooded pixels**. Despite this shortcoming, the synthetic image corresponds very well to the original SAR

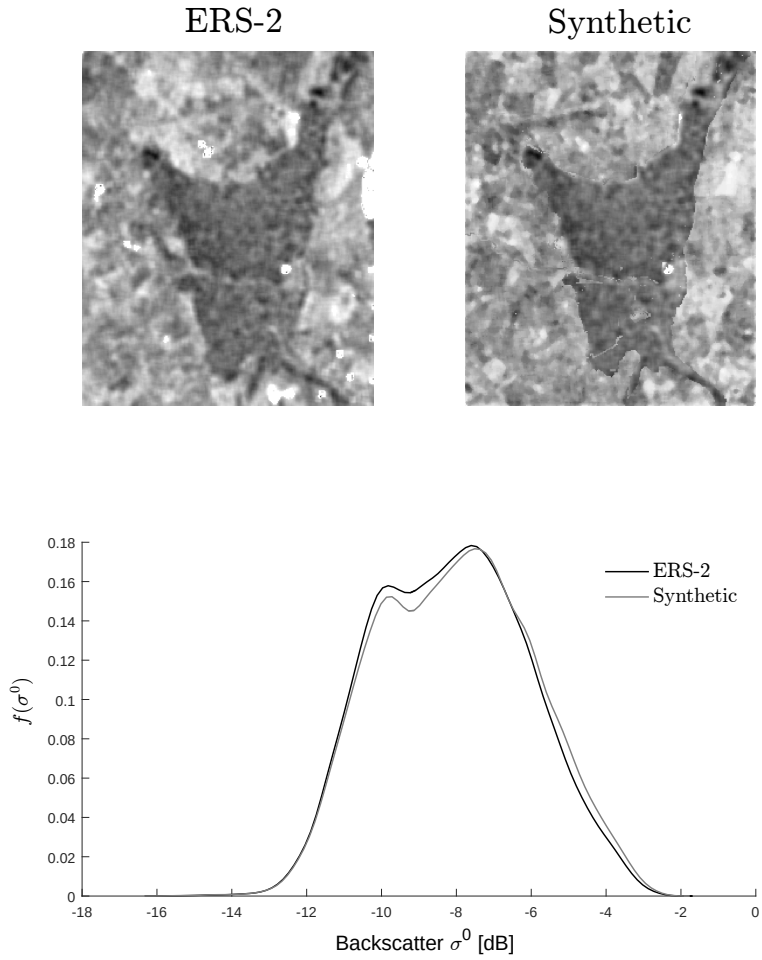


Figure 4: Illustration of the ERS-2 and synthetic image (12/12/06 11:00 am) (upper panel) and the pdf of the backscatter value for each image (lower panel).

image, and, as will be demonstrated, provides very similar results compared to those obtained from the original SAR image. Therefore, this framework can be used for generating SAR images at other points in time during the flood event.

3. Results and Discussion

3.1. Calibration with the ERS-2 SAR image

The calibration of the LISFLOOD-FP model with the ERS-2 image is performed through the use of the ROC-based calibration method. As stated before, the calibration is performed over 119 different sets of channel and floodplain

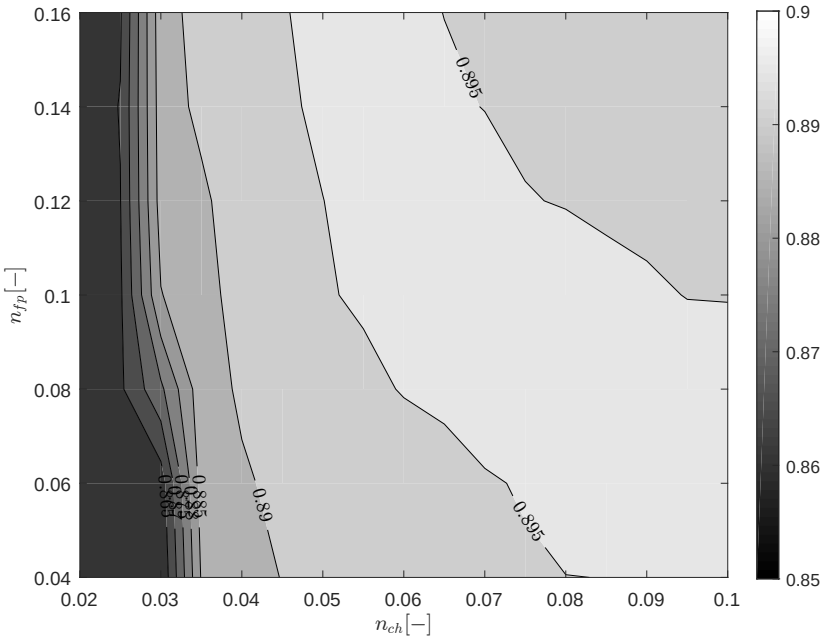


Figure 5: 2D calibration surface of the performance (AUC) for the channel (n_{ch}) and floodplain (n_{fp}) Manning coefficients. The AUC value ranges from 0 to 1. The best performance is observed for a channel and floodplain Manning coefficient of respectively 0.10 and 0.04.

Manning coefficients, which are placed regularly on a grid in the $n_{ch}n_{fp}$-plane. Results of this calibration are plotted in Figure 5. Visual inspection indicates a large degree of insensitivity of the model output to the ~~channel and~~ floodplain Manning coefficient **and a rather limited degree of insensitivity of the model output to the channel Manning coefficient**. The performance is more sensitive to lower values of the channel Manning coefficient; however, this sensitivity decreases for higher channel Manning coefficient values. The plot also shows that the model output is rather insensitive to the floodplain Manning coefficient, which can be explained by the low floodplain velocities causing minor energy losses.

The best AUC, equal to 0.8977, is obtained for a channel and floodplain Manning coefficient of respectively 0.10 and 0.04. It is important to note that these results are counter-intuitive, since one would expect a lower value for the channel Manning coefficient than what is found through calibration. Following the indicative values of Chow et al. (1988), the Manning coefficient for a natural channel is expected to vary between 0.02 and 0.05. However, Hunter et al. (2007)

state that values can exceed this range, because the parameters need to compensate for more than only the physical representation of the roughness conditions (for example, a simplified description of the hydraulic processes and geometry within the model). This could be a valid explanation, however, the value of 0.10 for the channel Manning coefficient remains an unexpectedly high value. Also, it is unexpected to find values for the channel Manning coefficient that are higher than the floodplain Manning coefficient.

Additionally, the performance of the best set (0.10, 0.04) is not significantly different from 87 parameter sets of the 118 other combinations (statistical test of DeLong et al. (1988) with Bonferoni-Holm correction). This thus leads to a typical equifinality problem where 88 parameter sets can be considered as behavioural, whereas the remaining 31 sets are then labelled as non-behavioural, since they lead to a significantly lower performance. The behavioural sets are covering the interval of 0.03 - 0.10 and 0.04 - 0.16 for the channel and floodplain Manning coefficient respectively. Yet, if other observational data were jointly available with the SAR image, this problem of equifinality could partly be resolved (Hunter et al., 2006).

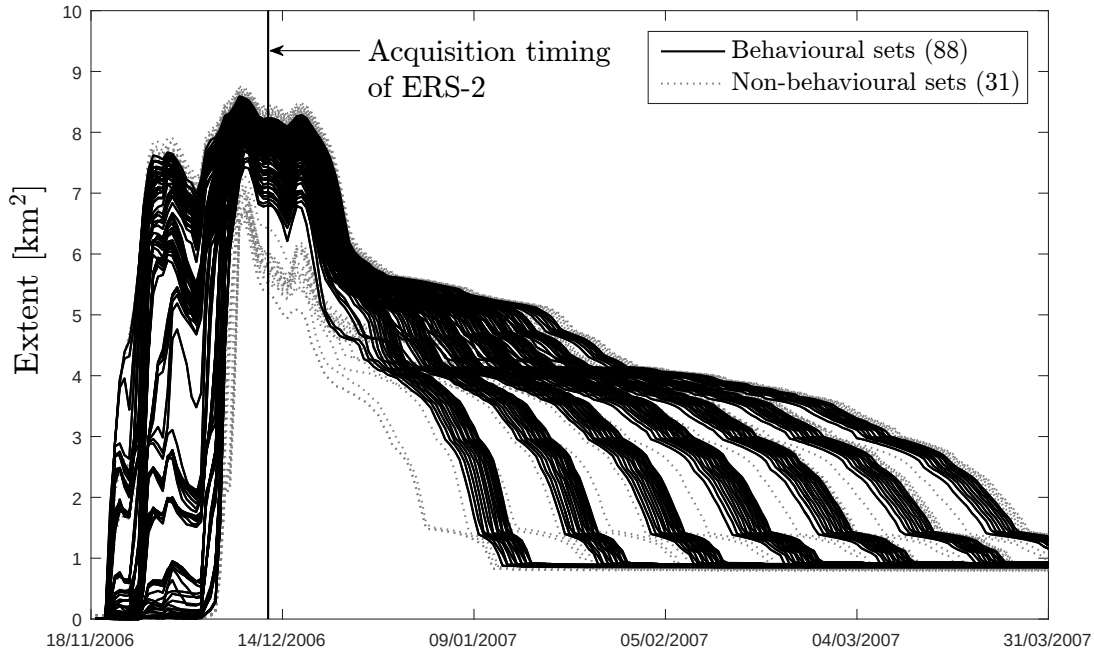
Parameter equifinality for the LISFLOOD-FP model, as revealed by Figure 5, has been previously reported in other case studies. Horritt et al. (2007) reported for the river Severn model that the calibration with extent-based performance measures based on images observing the flood at higher flows leads to a higher degree of parameter equifinality than when calibrated with images observing the flood at lower flows. Since the ERS-2 image observed the River Dee flood, just after peak flow, one could question whether the same issue is encountered here. More importantly, one would expect that an image observing the flood on the hydrograph rising (falling) limb would be more informative, since the image can give an indication about when the floodplain started to inundate (the water in the floodplain recedes). Therefore, the LISFLOOD-FP model outputs corresponding to the different Manning coefficient combinations are investigated as a function of time.

The modelled flood extents (km^2) are plotted for all parameter sets (be-

havioural and non-behavioural) as a function of time in Figure 6. It is important to note that the spread of the simulations is not constant in time. The plot shows the largest standard deviation in modelled extent between November 24 and December 6, a few days before flood peak occurs. During the time the maximum flood extent has been reached, the standard deviation remains nearly constant. However, once the recession of the flood starts, the standard deviation becomes larger again. Nevertheless, during recession, it does not reach the rising limb's level. Consequently, it can be hypothesized that calibration by means of flood observations obtained from the rising limb would lead to the identification of a more limited number of behavioural parameter sets. In the next section, this hypothesis will be investigated by calibrating the flood inundation model with several synthetic images taken at different points in time.

3.2. Calibration with synthetic images

The results in Section 3.1 clearly indicate a different temporal behaviour in the model predictions for the different behavioural parameter sets. In this section, we test whether this difference in model predictions depends on the timing of the acquisition of the SAR image that is used for calibration. In order to do this, a sequence of consistent SAR images is required. These images are generated by the synthetic framework presented in Section 2.4, using the backscatter data of both the image that was taken 3 days after the peak flow registered at Holt bridge and the dry reference image. The true state is generated by defining a parameter set ($\theta_{opt,true}$) and propagating a perturbation of the flow time series at Holt bridge through the LISFLOOD-FP model. Considering the results of the calibration in Section 3.1, indicating that the values of the true parameter set should be located in the interval [0.03-0.10] for the channel and [0.04-0.16] for the floodplain Manning coefficient, the true parameter set is defined as (0.06,0.12). The generated true state is used to create observations at 8 points in time, ranging from November 21 until March 14 (each time at 11:00 am), covering enough points in time to perform analysis before, during and after peak flow. The calibration exercise is hence performed 8 times, where



Date	21/11	27/11	12/12	21/12	02/01	16/01	09/02	14/03
st. dev. (km^2)	1.00	3.15	0.80	0.92	0.69	1.31	1.32	0.62

Figure 6: Modelled extent (km^2) as a function of time with non-behavioural and behavioural parameter sets. Each indicated point in time on the abscissa is given at 11:00 am. The standard deviation (st. dev.) of the modelled extents at several points in time is also indicated in the table below the plot.

each synthetic observation is used separately to identify the behavioural parameter sets based on the statistical test of DeLong et al. (1988). The results of the synthetic calibration for 8 points in time (code **A** to **H**) are shown in Figure 7. The summarized results are shown in Table 1.

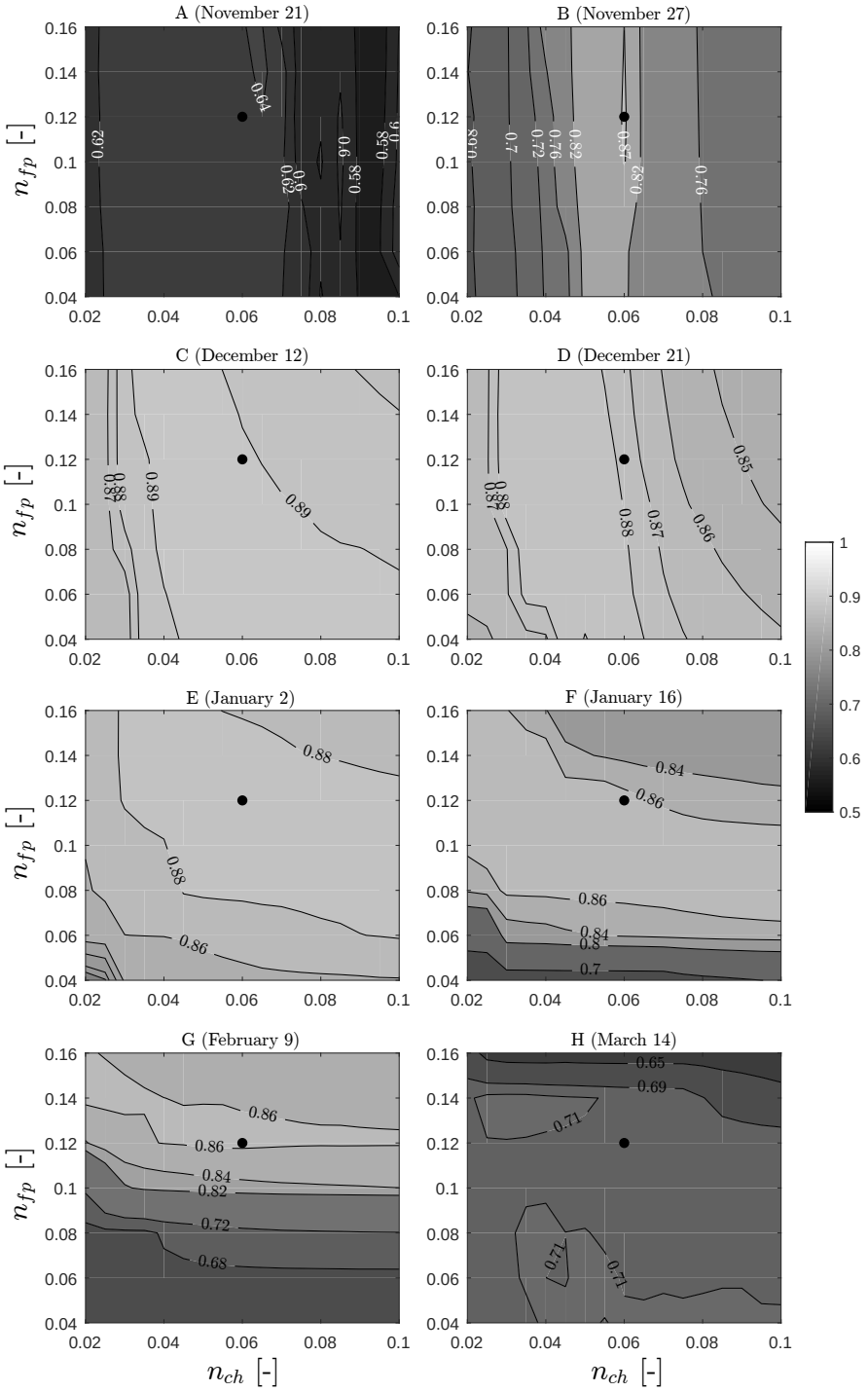


Figure 7: The 2D calibration surface of the performance (AUC) for the channel (n_{ch}) and floodplain (n_{fp}) Manning coefficients for 8 points in time (**A**, November 21 until **H**, March 14). The true parameter set $\theta_{true} = (0.06, 0.12)$ is indicated by a black dot.

Table 1: Summary statistics of the calibration. Every row summarizes the results for one of the 8 calibrations with the synthetic observation available on November 21 until March 14 (**A** to **H**). The flow is given at Holt bridge. N is the number of behavioural parameter sets.

	Flow [m^3/s]	**AUC**				***N***	**Interval behavioural sets**	
		Mean	Std	Min	Max	Total:119	n_{ch} channel roughness [-]	n_{fp} floodplain roughness [-]
A	75	0.611	0.019	0.568	0.644	93	0.02 - 0.10	0.04-0.16
B	101	0.762	0.053	0.669	0.877	3	0.06 - 0.06	0.10-0.14
C	160	0.886	0.009	0.863	0.894	88	0.03 - 0.10	0.04-0.16
D	100	0.870	0.014	0.843	0.890	15	0.03 - 0.075	0.04-0.16
E	30	0.874	0.017	0.777	0.886	75	0.03 - 0.10	0.06-0.16
F	30	0.826	0.068	0.659	0.877	50	0.02 - 0.10	0.08-0.16
G	30	0.777	0.086	0.666	0.864	49	0.02 - 0.10	0.12-0.16
H	30	0.697	0.027	0.632	0.717	98	0.02 - 0.10	0.04-0.14

First, the result of the calibrations shown in Figure 7 (**C**) is very similar to that in Figure 5, obtained when using the actual ERS-2 SAR image. Although the AUC-values are higher in Figure 5, a similar pattern in both figures is found, which for both cases results in an equal number of behavioural parameter sets (88). This result demonstrates the validity of the synthetic framework for assessing the impact of the timing of a SAR image acquisition of a flood on the calibration of the flood inundation model.

Visual inspection of the 8 calibration surfaces shown in Figure 7 indicates that, generally speaking, the sensitivity of the model output is at its highest when the model is calibrated using observations taken halfway through the rising limb, on November 27 (**B**). Here, the calibration problem is narrowed down to a small region covering only few parameter sets reaching AUC values that indicate clear optima in comparison to all other parameter sets. This small elongated region indicates a strong sensitivity to the channel Manning coefficient and insensitivity to the floodplain Manning coefficient. This insensitivity of the model to the floodplain Manning coefficient is also observed in Section 3.1 and in case studies of several authors (e.g. Bates et al. (2004)). On February 9 (**G**), during the recession of the flood water, the opposite is seen; the model now shows insensitivity to the channel Manning coefficient while also a narrow region covers optimal parameter sets, although it is less apparent. For all plots in between (**C**)-(**F**), there exists an intermediate state. Areas of optimal parameters are much broader, and the insensitivities slowly shift from the channel Manning coefficient to the floodplain Manning coefficient as the flood event is proceeding.

Table 1 confirms that the AUC values show a lower spread among the different parameter sets if pre-flood peak images are used in the calibration. In contrast, an increased number of behavioural sets is observed during peak flow, thus making it more difficult to identify the true set. For later points in time, the spread of behavioural sets becomes smaller again, more specifically in favour of the floodplain Manning coefficient. These results suggest that the ability to constrain the number of behavioural sets is lower when calibrated with an image observing the flood near peak flow. When using pre-flood peak images, a

smaller range of acceptable channel Manning coefficient values is found (column 8). The floodplain Manning coefficient is, however, more constrained when using post-flood images (column 9). It should be stated that the true set $\theta_{opt,true}$ (0.06,0.12) is included in the intervals of the behavioural sets for all points in time.

The reason for the insensitivity to channel and floodplain Manning coefficients is firstly related to the 'valley-filling' effect, often encountered in floodplains with relatively steep side slopes. At floodplain edges, when extent-based performance measures are used, a high increase in water height will only result in a small increase in the flooded area, *i.e.* the flood extent becomes more insensitive to flow changes. Since the event analysed here is of such a magnitude that it fills the floodplain completely for every proposed parameter set, the modelled extents around peak flow are relatively insensitive to the parameter values for the floodplain Manning coefficient. **However, it should also be noted that inflow hydrograph and flood extent graphs peak around the same point in time and follow a very similar shape, indicating that the floodplain fill does not explicitly affect the sensitivities**. Secondly, when low velocities are present in the floodplain, insensitivity to the floodplain Manning coefficient occurs (Di Baldassarre et al., 2009b). Since frictional drag F_D is proportional to the square of the water velocity v and a frictional drag coefficient η, related to the Manning coefficient n_{fp} ($F_D \sim \eta v^2$), small water velocities result in very low frictional drag, despite large values of η and thus n_{fp}. Lastly, the insensitivity is shifted to the channel Manning coefficient when the flood is receding. In that case, the roughness of the floodplain is a main actor in routing the water back to the river when boundary flow conditions have lost their significance towards river overflow.

3.3. Analysis of calibration results

One could wonder whether there are parameter sets that are behavioural irrespective of the data used for calibration. Therefore, this section presents an analysis of the results of the 8 calibrations simultaneously. The analysis should

indicate which parameter values yield the best model performance, irrespective of the timing of the SAR image acquisition. Therefore, we seek a set with the highest median performance and a small deviation in this performance. Figure 8 summarizes the results of each calibration in 4 plots. In the top panel, the median performance of the 8 calibration surfaces is plotted (left plot) with the difference in maximum and minimum performance of the 8 calibration surfaces (right plot). Additionally, in the bottom panel, the rank of the performances - going from 1 (worst) to 119 (best performing) is shown. Generally speaking, the four plots indicate that the channel and floodplain Manning coefficients, respectively ranging in [0.04-0.06] and [0.10-0.14], give the best performance for all the considered calibrations. For the channel Manning coefficient, these values reflect what is expected from Chow et al. (1988) however, for the floodplain Manning coefficient, values are consistently higher than expected which can be attributed reasons discussed above. These obtained optimal parameter ranges will thus always be identified as behavioural in the calibration procedure given any of the proposed data sets.

3.4. Impact of the timing of a SAR image acquisition on the uncertainty of the model-simulated flood extent

Results in Section 3.3 show that the modelled extent during the filling phase of the floodplain is very sensitive to the model parameter values (Figure 6). Therefore, one would expect that a larger range in the values of the behavioural parameter sets would lead to larger uncertainties in the predictions of the extent. The results of the synthetic calibration with images obtained on November 27 (**B**), December 12 (**C**) and February 9 (**G**) are used to assess the impact of the timing of the SAR image acquisition on the uncertainty of the modelled extent. For November 27, December 12 and February 9, respectively 3, 88 and 49 sets were identified as behavioural. Performing the calibration with the synthetic observations of November 27, December 12 and February 9 is referred to as respectively *scenario I, scenario II* and *scenario III*. The simulations of the extent of the 3, 88 and 49 behavioural sets are respectively shown in the

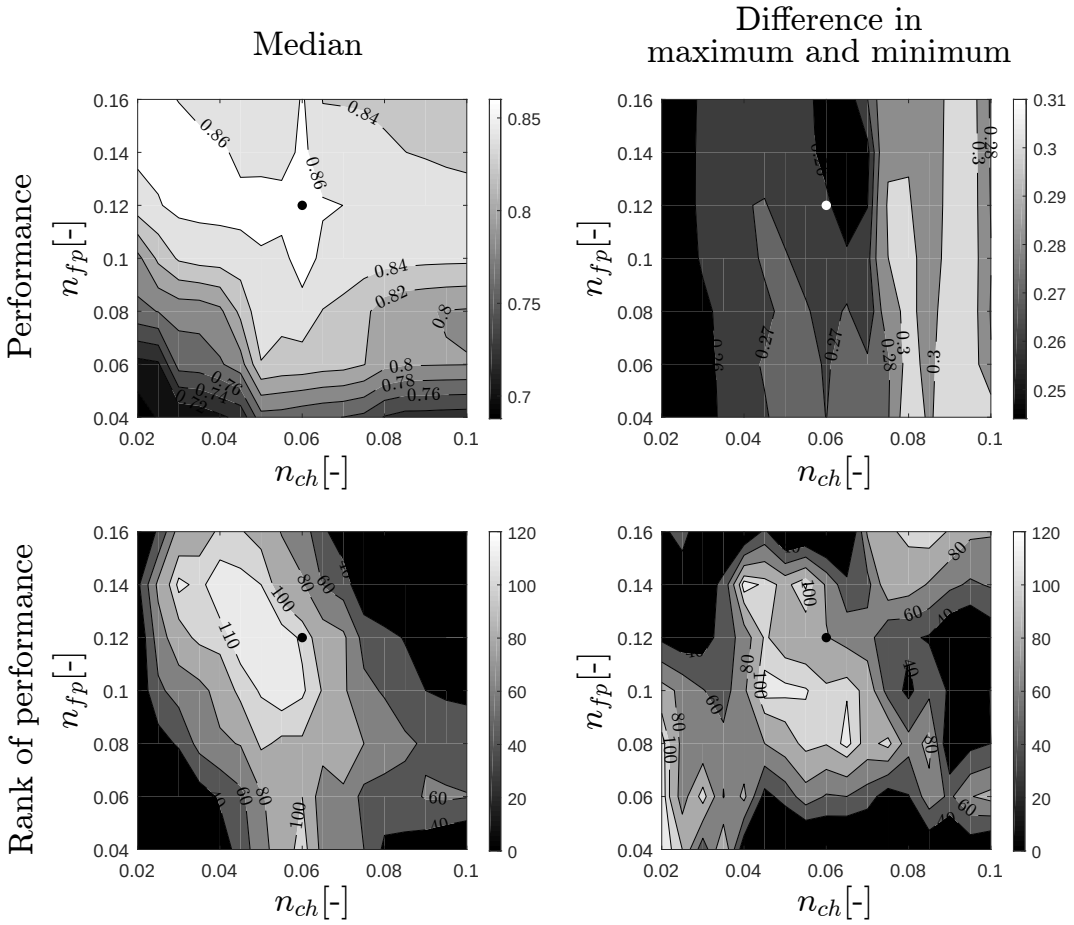


Figure 8: Summary of 8 calibration surfaces plotted in Figure 7. In the top panel, the median of the 8 performance values is calculated for every proposed set (left). Also the difference in the maximum and minimum in AUC is shown in the right plot. The bottom panel illustrate the same, but for the rank of the performance. The highest (lowest) rank is the best (worst) performing parameter set. The true parameter set $\theta_{true} = (0.06, 0.12)$ is indicated by a black or white dot.

top, middle and bottom panel of Figure 9. This figure reveals that, when calibration is performed with an image obtained before peak flow, the uncertainty of modelled extents increases along the receding limb. Similarly, for an image obtained after peak flow, uncertainty is large along the rising limb. This is again caused by the model output experiencing a changing sensitivity from the channel Manning coefficient to the floodplain Manning coefficient.

Figure 10 shows the maximum error at point in time t with respect to the flood extent for all three scenarios, calculated as follows:

$$\epsilon_{t,max} = \max\left(\left\{|E_{t,ts} - E_{t,sim,\theta_{opt,i}}| \;\middle|\; i \in \{1,..,N\}\right\}\right) \tag{3}$$

with $E_{t,ts}$ (km^2) the extent of the true state at point in time t, $E_{t,sim,\theta_{opt,i}}$ (km^2) the extent of simulation for the behavioural parameter set $\theta_{opt,i}$ at point in time t and N the total number of behavioural parameter sets. As discussed above, the largest error for *scenarios II* and *III* is observed before the peak flow. For the full simulation time, the error reaches a maximum of 1.51 km^2 for *scenario I*, while for *scenario II* and *scenario III*, the maximum value is very similar, respectively 6.36 (km^2) and 6.54 (km^2). On December 12 (**C**), an error of 0.09 km^2, 1.20 km^2 and 2.21 km^2 is respectively observed for *scenario I*, *scenario II* and *scenario III*. It is clear that the calibration results have a large impact on the error of the modelled flood extent. If one would calibrate the model with an image observing the flood before peak flow, the errors on the modelled extents would be smaller than when an image observing the flood during or after peak flow would be used for the calibration.

4. Conclusions

This study aimed at investigating whether the timing of the acquisition of a SAR image within a flood event has an important impact on the calibration of a flood inundation model. For the considered river Dee flood (December 2006), one SAR image was available as well as a dry reference SAR image (August 2006), which allowed to generate images through a synthetic framework.

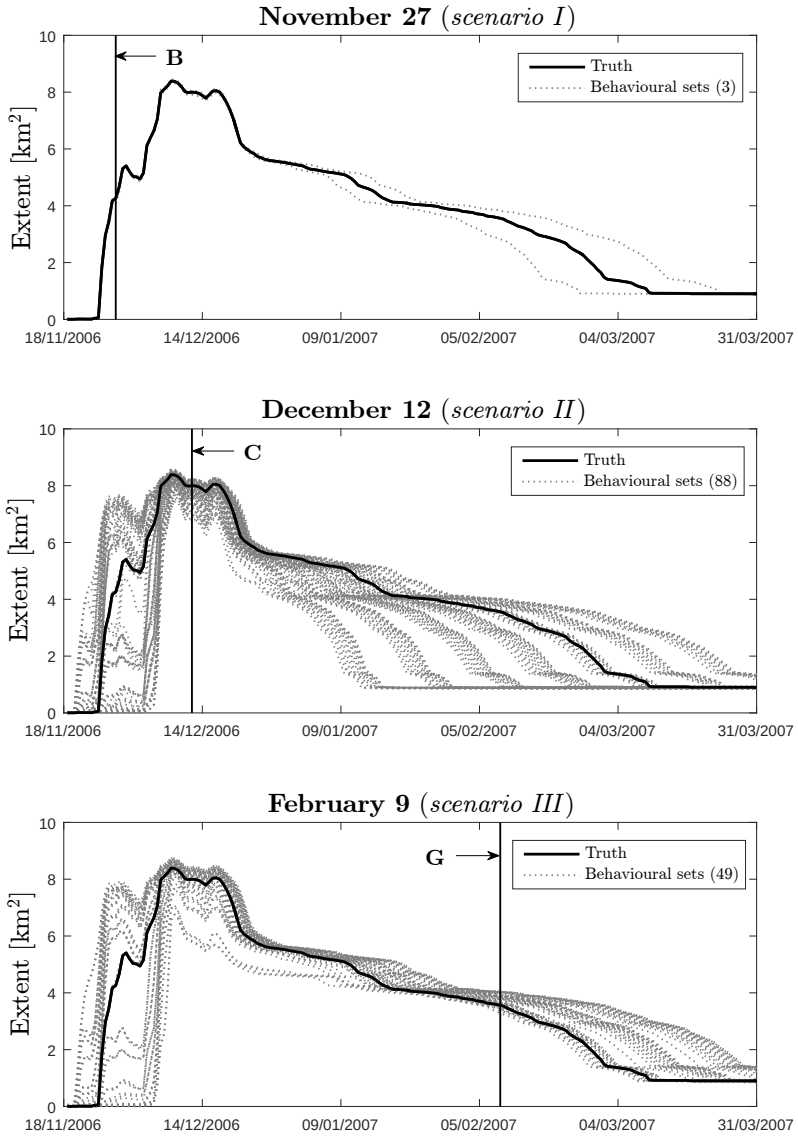


Figure 9: Plot of the flood extent for the behavioural parameter sets determined with the synthetic observations obtained on November 27 (*scenario I*), December 6 (*scenario II*) and February 9 (*scenario III*). For each scenario, the simulated extents are plotted corresponding to respectively 3, 88 and 49 obtained behavioural parameter sets. The true extent (Truth) is also shown.

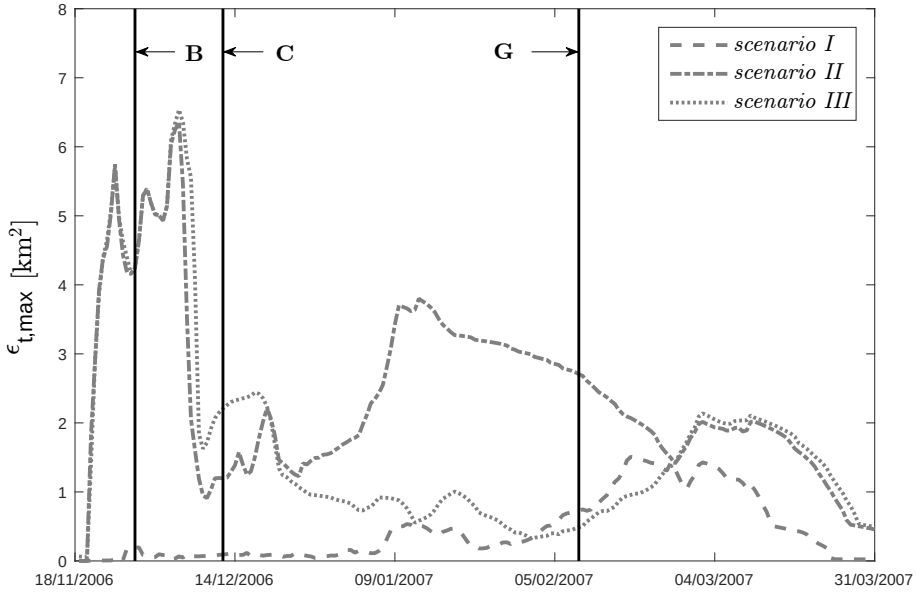


Figure 10: Maximum error $\epsilon_{t,max}$ for *scenarios I, II* and *III*. Points in time for the synthetic observations on November 27 (**B**), December 6 (**C**) and February 9 (**G**) are also indicated.

The channel and floodplain Manning coefficients of the LISFLOOD-FP inundation model were calibrated with a real ERS-2 SAR image observing the flood near the peak flow of the flood event. The calibration resulted in a high degree of equifinality, leading to a large number of behavioural parameter sets, for which it was concluded that the flood extent observed by the SAR image is insufficiently informative to constrain the parameter space. In a second experiment, the model was calibrated several times, each time with a different synthetic SAR image of the flood at a different point in time. The results indicated that it is more beneficial to constrain the parameter space when the model is calibrated with an image observing the flood during the rising limb of the hydrograph. Calibration performed with an image taken during peak flow or during the receding limb is unable to properly constrain the parameter space. The preference for pre-peak observations should be sought in the fact that during the rising phase, the velocity of the water is mainly influenced by the roughness of the channel and to a lesser extent also by the roughness of the floodplain. At a certain water height, the water level in this particular floodplain may still increase whereas the spatial extent does not significantly change any more (valley-filling effect). Once this situation is reached, calibration with a SAR image leads to a larger equifinality in roughness parameters. During the initial recession, water is released without a major change in flood extent (as for the full valley, a decrease in water volume does not cause a significant change in flood extent). Hereafter, the routing of water is mainly supported by both floodplain Manning coefficient and by downstream conditions, and is less affected by the roughness of the channel.

Since flood risk management requires an accurate estimate of the flood extent, an accurately calibrated model is **useful**, and uncertainty of model predictions due to parameter uncertainty is not desirable. To assess this, the uncertainty of the predicted extent due to the equifinality in parameter sets that resulted from the calibration with images of each different acquisition timing was also tested. Therefore, the modelled extents for the behavioural parameter sets determined through calibration with images at 3 different points in time

(before, during and after peak flow), were compared with each other. These results showed that the maximum error on the simulated extent is consistently smaller when an image acquired before the flood peak is used for calibrating the model.

Satellite SAR data have a great potential to support the calibration of flood inundation models. However, it was shown that the time of the acquisition with respect to the flood peak is of paramount importance, since pre-flood peak images allow for a better constraining of the model parameter space, and a reduced model prediction uncertainty than when post-flood peak images are used in the calibration. Unfortunately, it is more difficult to obtain images at the rising limb of the hydrograph as this is typically shorter, but also because some satellites require targeting and this tends to be done once a major flood has happened as the image can then be used for disaster management.

5. Acknowledgements

The work in this paper has been funded by the Belgian Science Policy for the FLOODMOIST project in the framework of the STEREO II programme and through project G.0179.16N of the Research Foundation Flanders (FWO) and project 01J04015 of the "Bijzonder Onderzoeksfonds" of Ghent University. Hans Lievens is a postdoctoral research fellow of the FWO. Part of Guy Schumann's time on this manuscript was spent at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, the Hercules Foundation and the Flemish Government department EWI.

References

Anderson, M.G., Walling, D.E., Bates, P.D.. Floodplain Processes. John Wiley, 1996.

Aronica, G.T., Bates, P.D., Horritt, M.S.. Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. Hydrological Processes 2002;16(10):2001–2016. doi:10.1002/hyp.398.

Aronica, G.T., Hankin, B., Beven, K.J.. Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data. Advances in Water Resources 1998;22(4):349–365. doi:10.1016/S0309-1708(98)00017-7.

Bates, P.D.. Remote sensing and flood inundation modelling. Hydrological Processes 2004;18(13):2593–2597. doi:10.1002/hyp.5649.

Bates, P.D., Anderson, M., Baird, L., Walling, D., Simm, D.. Modelling floodplain flows using a two-dimensional finite element model. Earth Surface Processes and Landforms 1992;17(6):575–588. doi:10.1002/esp.3290170604.

Bates, P.D., De Roo, A.. A simple raster-based model for flood inundation simulation. Journal of Hydrology 2000;236(1-2):54–77. doi:10.1016/S0022-1694(00)00278-X.

Bates, P.D., Horritt, M.S., Aronica, G., Beven, K.. Bayesian updating of flood inundation likelihoods conditioned on flood extent data. Hydrological Processes 2004;18(17):3347–3370. doi:10.1002/hyp.1499.

Bates, P.D., Horritt, M.S., Fewtrell, T.J.. A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. Journal of Hydrology 2010;387(1-2):33–45. doi:10.1016/j.jhydrol.2010.03.027.

Beven, K.J.. A discussion of distributed hydrological modelling. In: Abbott, M.B., Refsgaard, J.C., editors. Distributed Hydrological Modelling. Springer Netherlands; volume 22 of *Water Science and Technology Library*; 1996. p. 255–278.

Beven, K.J., Binley, A.. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes 1992;6(3):279–298. doi:10.1002/hyp.3360060305.

Bradley, A.P.. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition 1997;30(7):1145–1159. doi:10.1016/S0031-3203(96)00142-2.

Chow, V.T., Maidment, D.R., Mays, L.W.. Applied Hydrology. New York: McGraww-Hill, 1988.

DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L.. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44(3):837–845. doi:10.2307/2531595.

Di Baldassarre, G.. Floods in a changing climate: inundation modeling. Cambridge University Press, 2012.

Di Baldassarre, G., Montanari, A.. Uncertainty in river discharge observations: a quantitative analysis. Hydrology and Earth System Sciences Discussions 2009;6(1):39–61. doi:10.5194/hessd-6-39-2009.

Di Baldassarre, G., Schumann, G.J.P., Bates, P.D.. A technique for the calibration of hydraulic models using uncertain satellite observations of flood extent. Journal of Hydrology 2009a;367(3-4):276–282. doi:10.1016/j.jhydrol.2009.01.020.

Di Baldassarre, G., Schumann, G.J.P., Bates, P.D.. Near real time satellite imagery to support and verify timely flood modelling. Hydrological Processes 2009b;23(85):799–803. doi:10.1002/hyp.7229.

Efron, B.. Bootstrap methods - Another look at the Jackknife. Annals of Statistics 1979;7(1):1–26. doi:10.1214/aos/1176344552.

Evensen, G.. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dynamics 2003;53(4):343–367. doi:10.1007/s10236-003-0036-9.

Fawcett, T.. An introduction to ROC analysis. Pattern Recognition Letters 2006;27(8):861–874. doi:10.1016/j.patrec.2005.10.010.

García-Pintado, J., Neal, J.C., Mason, D.C., Dance, S.L., Bates, P.D.. Scheduling satellite-based SAR acquisition for sequential assimilation of water level observations into flood modelling. Journal of Hydrology 2013;495:252–266. doi:10.1016/j.jhydrol.2013.03.050.

Giustarini, L., Hostache, R., Matgen, P., Schumann, G.J.P., Bates, P.D., Mason, D.C.. A change detection approach to flood mapping in urban areas using TerraSAR-X. IEEE Transactions on Geoscience and Remote Sensing 2013;51(4):2417–2430. doi:10.1109/TGRS.2012.2210901.

Giustarini, L., Vernieuwe, H., Verwaeren, J., Chini, M., Hostache, R., Matgen, P., Verhoest, N.E.C., De Baets, B.. Accounting for image uncertainty in SAR-based flood mapping. International Journal of Applied Earth Observation and Geoinformation 2015;34:70–77. doi:10.1016/j.jag.2014.06.017.

Horritt, M.S.. A statistical active contour model for SAR image segmentation. Image and Vision Computing 1999;17(3-4):213–224. doi:10.1016/S0262-8856(98)00101-2.

Horritt, M.S., Bates, P.D.. Effects of spatial resolution on a raster based model of flood flow. Journal of Hydrology 2001;253:239–249. doi:10.1016/S0022-1694(01)00490-5.

Horritt, M.S., Bates, P.D.. Evaluation of 1D and 2D numerical models for predicting river flood inundation. Journal of Hydrology 2002;268(1-4):87–99. doi:10.1016/S0022-1694(02)00121-X.

Horritt, M.S., Di Baldassarre, G., Bates, P.D., Brath, A.. Comparing the performance of a 2D finite element and a 2D finite volume model of floodplain inundation using airborne SAR imagery. Hydrological Processes 2007;21(20):2745–2759. doi:10.1002/hyp.6486.

Hostache, R., Matgen, P., Schumann, G.P., Puech, C., Hoffmann, L., Pfister, L.. Water level estimation and reduction of hydraulic model calibration uncertainties using satellite SAR images of floods. IEEE Transactions on Geoscience and Remote Sensing 2009;47(2):431–441. doi:10.1109/TGRS.2008.2008718.

Hunter, N.M., Bates, P.D., Horritt, M.S., De Roo, A., Werner, M.G.F.. Utility of different data types for calibrating flood inundation models within a GLUE framework. Hydrology and Earth System Sciences 2005;9(4):412–430. doi:10.5194/hess-9-412-2005.

Hunter, N.M., Bates, P.D., Horritt, M.S., Wilson, M.D.. Improved simulation of flood flows using storage cell models. Proceedings of the Institute of Civil Engineers, Water Management 2006;159(1):9–18. doi:10.1680/wama.2006.159.1.9.

Hunter, N.M., Bates, P.D., Horritt, M.S., Wilson, M.D.. Simple spatially-distributed models for predicting flood inundation: A review. Geomorphology 2007;90(3-4):208–225. doi:10.1016/j.geomorph.2006.10.021.

Lang, M., Pobanz, K., Renard, B., Renouf, E., Sauquet, E.. Extrapolation of rating curves by hydraulic modelling, with application to flood frequency analysis. Hydrological Sciences Journal 2010;55(6):883–898. doi:10.1080/02626667.2010.504186.

Mason, D.C., Horritt, M.S., Dall'Amico, J.T., Scott, T.R., Bates, P.D.. Improving river flood extent delineation from synthetic aperture radar using airborne laser altimetry. IEEE Transactions on Geoscience and Remote Sensing 2007;45(12):3932–3943. doi:10.1109/TGRS.2007.901032.

Matgen, P., Hostache, R., Schumann, G., Pfister, L., Hoffmann, L., Savenije, H.. Towards an automated SAR-based flood monitoring system: Lessons learned from two case studies. Physics and Chemistry of the Earth, Parts A/B/C 2011;36(7-8):241–252. doi:10.1016/j.pce.2010.12.009.

Matgen, P., Montanari, M., Hostache, R., Pfister, L., Hoffmann, L., Plaza, D., Pauwels, V.R.N., De Lannoy, G.J.M., De Keyser, R., Savenije, H.. Towards the sequential assimilation of SAR-derived water stages into hydraulic models using the Particle Filter: proof of concept. Hydrology and Earth System Sciences 2010;14(9):1773–1785. doi:10.5194/hess-14-1773-2010.

Montanari, M., Hostache, R., Matgen, P., Schumann, G.J.P., Pfister, L., Hoffmann, L.. Calibration and sequential updating of a coupled hydrologic-hydraulic model using remote sensing-derived water stages. Hydrology and Earth System Sciences 2009;13(3):367–380. doi:10.5194/hess-13-367-2009.

Neal, J., Keef, C., Bates, P.D., Beven, K., Leedal, D.. Probabilistic flood risk mapping including spatial dependence. Hydrological Processes 2013;27(9):1349–1363. doi:10.1002/hyp.9572.

Neal, J.C., Bates, P.D., Fewtrell, T.J., Hunter, N.M., Wilson, M.D., Horritt, M.S.. Distributed whole city water level measurements from the Carlisle 2005 urban flood event and comparison with hydraulic model simulations. Journal of Hydrology 2009;368(1-4):42–55. doi:10.1016/j.jhydrol.2009.01.026.

Neal, J.C., Schumann, G.J.P., Bates, P.D.. A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. Water Resources Research 2012;48(11):W11506. doi:10.1029/2012WR012514.

Pappenberger, F., Beven, K., Horritt, M.S., Blazkova, S.. Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations. Journal of Hydrology 2005;302(1–4):46–69. doi:10.1016/j.jhydrol.2004.06.036.

Pappenberger, F., Matgen, P., Beven, K.J., Henry, J.B., Pfister, L., de Fraipont, P.. Influence of uncertain boundary conditions and model structure on flood inundation predictions. Advances in Water Resources 2006;29(10):1430–1449. doi:10.1016/j.advwatres.2005.11.012.

Romanowicz, R., Beven, K., Tawn, J.. Bayesian calibration of flood inundation models. In: M.G. Anderson, D.W., Bates, P., editors. Flood Plain Processes. Chichester: John Wiley and Sons; 1996. p. 181–196.

Schumann, G.J.P., Bates, P.D., Horritt, M.S., Matgen, P., Pappenberger, F.. Progress in integration of remote sensing-derived flood extent and stage data and hydraulic models. Reviews of Geophysics 2009a;47(4):1–20. doi:10.1029/2008RG000274.

Schumann, G.J.P., Di Baldassarre, G., Bates, P.D.. The utility of spaceborne radar to render flood inundation maps based on multialgorithm ensembles. IEEE Transactions on Geoscience and Remote Sensing 2009b;47(8):2801–2807. doi:10.1109/TGRS.2009.2017937.

Schumann, G.J.P., Vernieuwe, H., De Baets, B., Verhoest, N.E.C.. ROC-based calibration of flood inundation models. Hydrological Processes 2014;28(22):5495–5502. doi:10.1002/hyp.10019.

Stephens, E., Bates, P.D., Freer, J.E., Mason, D.C.. The impact of uncertainty in satellite data on the assessment of flood inundation models. Journal of Hydrology 2012;414-415:162–173. doi:10.1016/j.jhydrol.2011.10.040.

Tarpanelli, A., Brocca, L., Melone, F., Moramarco, T.. Hydraulic modelling calibration in small rivers by using coarse resolution synthetic aperture radar imagery. Hydrological Processes 2012;27(9):1321–1330. doi:10.1002/hyp.9550.

Werner, M., Hunter, N.M., Bates, P.D.. Identifiability of distributed floodplain roughness values in flood extent estimation. Journal of Hydrology 2005;314(1-4):139–157. doi:10.1016/j.jhydrol.2005.03.012.